



August 1, 2018

The Research Says – Or Perhaps It Doesn't?

Alanna Bjorklund-Young, Senior Research and Policy Analyst

David Steiner, Executive Director

Earlier this year, the American Enterprise Institute (AEI) published a new study,¹ by Michael McShane, Patrick Wolf, and Collin Hitt, about the predictive power of test scores for long-term outcomes among students in schools of choice. The study claims to be “a meta-analysis on the effect that school choice has on educational attainment and [to show] that, at least for school choice programs, there is a weak relationship between impacts on test scores and later attainment outcomes.” The authors conclude, “Policymakers need to be much more humble in what they believe that test scores tell them about the performance of schools of choice: Test scores should not automatically occupy a privileged place over parental demand and satisfaction as short-term measures of school choice success or failure.”

The study and its provocative findings received significant interest, including a week's worth of daily critiques² from Michael Petrilli of the Thomas B. Fordham Institute, a negative review from Christopher Lubienski of Indiana University and T. Jameson Brewer of the University of North Georgia,³ and support from Corey A. DeAngelis of the Cato Institute.⁴ Most of the critiques and support focus on the study's methodology—discussing, for example, if the researchers use the correct definition of school choice, did the authors correctly report studies' findings, and were the correct meta-analysis methods used.

¹ Michael McShane, Patrick Wolf, and Collin Hitt, “Do Impacts on Test Scores Even Matter? Lessons from Long-Run Outcomes in School Choice Research: Attainment Versus Achievement Impacts and Rethinking How to Evaluate School Choice Programs” (American Enterprise Institute, March 19, 2018), <http://www.aei.org/publication/do-impacts-on-test-scores-even-matter-lessons-from-long-run-outcomes-in-school-choice-research/>.

² “How to Think about Short-Term Test Score Changes and Long-Term Student Outcomes | The Thomas B. Fordham Institute,” <https://edexcellence.net/articles/how-to-think-about-short-term-test-score-changes-and-long-term-student-outcomes>; “When Looking Only at School Choice Programs, Both Short-Term Test Scores and Long-Term Outcomes Are Overwhelmingly Positive | The Thomas B. Fordham Institute,” <https://edexcellence.net/articles/when-looking-only-at-school-choice-programs-both-short-term-test-scores-and-long-term>; “For the Vast Majority of School Choice Studies, Short- and Long-Term Impacts Point in the Same Direction | The Thomas B. Fordham Institute,” <https://edexcellence.net/articles/for-the-vast-majority-of-school-choice-studies-short-and-long-term-impacts-point-in-the>; “Findings about School Choice Programs Shouldn't Be Applied to Individual Schools | The Thomas B. Fordham Institute,” <https://edexcellence.net/articles/findings-about-school-choice-programs-shouldnt-be-applied-to-individual-schools>; “Yes, Impacts on Test Scores Matter | The Thomas B. Fordham Institute,” <https://edexcellence.net/articles/yes-impacts-on-test-scores-matter>.

³ Christopher Lubienski and T. Jameson Brewer, “NEPC Review: Do Impacts on Test Scores Even Matter? (American Enterprise Institute, March 2018),” *National Education Policy Center*, May 3, 2018, <https://nepc.colorado.edu/thinktank/review-goalposts>.

⁴ Corey A. DeAngelis This article appeared in *The 74 Million* on April 26 and 2018., “5 Reasons Why 5-Point Takedown of AEI Report on School Choice, Tests, and Long-Term Outcomes Misses the Mark,” Cato Institute, April 26, 2018, <https://www.cato.org/publications/commentary/5-reasons-why-5-point-takedown-aei-report-school-choice-tests-long-term>.

While it is important to raise relevant methodological issues, we will argue that even if the study had met the highest standards of methodological rigor, the research findings would not adequately support the conclusions the authors claim. We believe that the study's logic, findings, and policy recommendations serve instead as a useful case study of the pitfalls of simply skimming over the text of research studies and focusing on the conclusion. It is exactly because of the critical role evidence *should* have in education policies that this tale is worth telling: a case study in why we need to slow down and ask tough questions, especially when faced with research findings that are not as definitive as they first appear.

Case Study: Evaluating Research Carefully

Most of the responses to the AEI report question the research report's methodology. Here, we do not contribute to that debate but instead assume for a moment that the methodology of this study were completely sound and rigorous. We still suggest three places that a critical reader should interrogate the report's assumptions and conclusions.

1. *Question the underlying assumptions.*

While the authors claim that this research is about school choice, they fail to show that their findings say anything specific about school choice. In order for these results to draw conclusions about school choice specifically, the authors would have needed first to conduct an initial analysis that shows that test scores reliably predict long-term outcomes among schools *in general*. Then the authors could place the research at hand in context, by showing that the relationship between test scores and long-term outcomes does not hold specifically for *schools of choice*. That kind of evidence would have allowed the authors to argue that accountability protocols for schools of choice should be different from those for assigned district schools. If the authors found initially that test scores could not be used to reliably predict long-term outcomes amongst any kind of school – assigned or chosen – then they would be making a different point altogether: one about the limits of relying on test data in K-12 education. The authors fail to establish their original premise.

Next, the authors assume that the value of test scores is only to signal long-term outcomes. The researchers state that “even the most fervent believer in the power of standardized tests agrees that test scores are merely an interim measure. There is no point in increasing test scores for test scores’ sake. Increased test scores are supposed to indicate progress toward more important long-term results.”⁵ Is this statement true for all grade levels? For example, are 3rd-grade ELA test scores (or elementary scores generally) merely an interim measure of a longer-term outcome, such as high-school graduation? Or, is a primary value of elementary test scores to provide a measure of a child’s basic reading (or math) skills, while also acknowledging that this is an important skill required for long-term outcomes, such as high-school graduation? In contrast, one might expect that some high school test scores, such as 10th grade math scores, are better predictors of long-term outcomes, such as high school graduation (e.g. high scores on these tests require both math skills and fairly strong study habits). Even here, however, we might argue that ensuring that students have some basic level of high school math proficiency is an important purpose of the test in itself, and that test scores matter in their own right, for this reason.

But even if we accepted the premise that the value of test scores lies in long-term outcomes, we should question the three they choose: high-school graduation, college enrollment, and college graduation. These

⁵ Page 1

three do not exhaust the outcomes of interest. Some policymakers and scholars would likely argue that the real purpose of school is to produce active citizens who are productive members of society. If so, then labor-market outcomes and contributions to society (through taxes, civic engagement, or at least not committing crimes) are the important outcomes to measure. Others might argue that social outcomes such as physical health and marriage are also important. Imagine if the researchers had found that test scores are strong predictors for such other long-term outcomes. In such circumstances, they might well have changed their conclusions.

2. *Question whether the evidence justifies the conclusions.*

The findings of this research are: “A school choice program’s impact on test scores is a weak predictor of its impacts on longer-term outcomes. Our findings are based on 39 unique impact estimates across studies of more than 20 programs.”⁶

The researchers conclude, “Our findings suggest that focusing on test scores may lead authorities to favor the wrong school choice programs. Focusing on test score gains may lead regulators to favor schools whose benefits could easily fade over time and punish schools that are producing long-lasting gains.”⁷

The researchers here betray their concern: that “choice” schools (as they very broadly define them) produce high long-term gains, but look weak when judged by test scores. Is this concern warranted? To be warranted, the researchers would have to have found that choice schools have a positive and statistically significant impact on students’ long-term outcomes, but a negative and statistically significant impact on students’ test scores. Did the authors find strong evidence of this? No. Out of 34 results that looked at the relationship between **ELA test scores** and high school graduation impacts, only one result showed negative and statistically significant test score results, but positive and statistically significant high school graduation results. Similarly, out of the 33 results that looked at the relationship between **math test scores** and high school graduation rates, only one result showed statistically significant negative math test scores and statistically significant positive results for high school graduation rates. *There were no such mismatches among test score results and the other long-term outcomes used by the researchers.* Therefore, their general conclusion appears to only be relevant to a very small sample of programs. Given the weakness in the evidence, therefore, should we really be concerned that authorities will “punish schools that are producing long-lasting gains”? The evidence provided certainly does not suggest so.

3. *Question the realism of the policy recommendations.*

Let’s review the researchers’ full conclusions: “The policy implications from this analysis are clear. The most obvious implication is that policymakers need to be much more humble in what they believe that test scores tell them about the performance of schools of choice. Test scores are not giving us the whole picture. Insofar as test scores are used to make determinations in ‘portfolio’ governance structures or are used to close (or expand) schools, policymakers might be making errors. This is not to say that test scores should be wholly

⁶ Page 2

⁷ Page 2

discarded. Rather, test scores should be put in context and should not automatically occupy a privileged place over parental demand and satisfaction as short-term measures of school choice success or failure.”⁸

It is not clear that anyone actually believes that test scores provide the entire picture of what children have learned in school. However, *which* measures should be used instead of, or in addition to, test scores is another matter. Under ESSA,⁹ measures of student learning used for accountability purposes must meet the following criteria: they must be discernible in the aggregate and also for student subgroups; be comparable across a state’s school districts; distinguish differences in performance among schools; be valid;¹⁰ be reliable;¹¹ and have a proven impact on student achievement. These requirements are important for ensuring that the measures are meaningful and actually provide data that is comparable across schools and school populations. Chronic absenteeism¹² and measures of school climate are two short-term measures that both meet the ESSA requirements and that many states have adopted in their accountability systems. Movement towards using multiple measures, each of which adds valid and reliable information about the school’s environment and impact on students certainly provides a more complete picture of school quality than test scores alone.

It is surprising, however, that the researchers recommend – as a policy change – that parental demand and parental satisfaction be assigned *more* weight than test scores, while not simultaneously providing strong justification for their use or a clear way to measure them. We would have needed a careful definition of what counts as parental demand and how to measure parental satisfaction.

More importantly, even if we were to accept (which we do not) the researchers’ argument that test scores are a poor measure of choice schools, the researchers would need to advocate for the importance of parental demand in the states’ accountability system on one of two bases: 1) that choice is a good in itself, and 2) that schools of choice, by contrast to non-schools of choice, produce better long-term outcomes. The first is an ideological claim independent of educational outcomes. The second is an empirical claim, which the researchers also leave unaddressed. If the researchers had been consistent in their logic, they would have to have shown that parental demand and satisfaction are stronger predictors of positive long-term outcomes than test scores. This is, after all, the argument the researchers used to advocate for the limited use of test scores.

⁸ Page 20

⁹ Every Student Succeeds Act, *Of 2015*, Pub (L, 2015).

¹⁰ I.e. the measure actually measures what you think it does.

¹¹ I.e. the measure is consistent—it produces similar results under similar conditions.

¹² Typically defined as missing 10% or more of school, including both excused and unexcused absences.